Proc. Natl. Acad. Sci. USA
Vol. 74, No. 10, pp. 4325–4328, October 1977
Biochemistry

Gene-sized DNA molecules of the *Oxytricha* macronucleus have the same terminal sequence

(palindromes/inverted terminal repeats/DNA sequence/ciliated protozoan)

RICHARD M. LAWN*

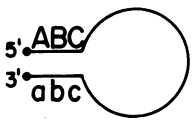
Department of Molecular, Cellular and Developmental Biology, University of Colorado, Boulder, Colorado 80309

Communicated by David M. Prescott, August 2, 1977

ABSTRACT **The DNA in the macronucleus of the ciliated protozoan *Oxytricha* exists as small linear molecules with a number average size of about 3000 base pairs. Most, and possibly all, of these DNA molecules contain the same inverted terminal repeat sequence of approximately 26 base pairs. In addition to its terminal location, two inverted copies of this same sequence surround single-strand interruptions within these DNA molecules. This sequence arrangement may function in the processing of these molecules from large chromosomal precursors or in the subsequent replication of these small linear DNAs during cell reproduction.**

Ciliates contain two types of nuclei, micronuclei and macronuclei. The micronucleus of *Oxytricha* is diploid and divides by mitosis. It apparently serves only as a germ line nucleus which undergoes meiosis and intercellular exchange during conjugation (1). Macronuclei, on the other hand, divide amitotically during vegetative growth of the organism. After conjugation, however, macronuclei form anew from micronuclear precursors. The conjugating pair of cells exchanges haploid micronuclei, and fusion occurs between an exchanged and a resident haploid micronucleus. All old nuclei degenerate, while the fusion nucleus undergoes two mitotic divisions before one of the daughters begins to develop into a new macronucleus (1). This development commences with DNA replication and the formation of polytene chromosomes which resemble those of dipteran salivary glands (2). The polytene chromosomes are then transected through the interbands, forming a large number of small fragments. Extensive DNA degradation follows, with selective elimination of as much as 97% of the micronuclear sequence complexity (3). A second round of replication ensues, yielding the mature macronucleus, which consists entirely of small DNA molecules ranging from about 550 to 30,000 base pairs (BP), with a number average size of approximately 3000 BP (4–6). About 1000 copies of each of the approximately 17,000 different DNA molecules exist in a macronucleus (3). The size of these DNAs implies the coding potential of one to a few genes per molecule. Yet this collection of small DNA molecules of greatly diminished sequence complexity directs all the vegetative growth of the organism. The macronuclear DNA serves as template for all the RNA synthesis (except mitochondrial) of *Oxytricha*, and cells lacking micronuclei can grow and divide extensively (4). Here I demonstrate the presence of a unique DNA sequence that is present in inverted order at each end of these molecules.

Wesley (5) noted that, upon denaturation and brief reannealing, most macronuclear DNAs formed single-stranded circles, presumably closed by short, complementary DNA sequences located at or near the ends of the strands. This is shown schematically below, where "ABC" and "abc" represent a DNA sequence and its complement.



He estimated the length of these duplex "neck" regions as <50 BP because they could not be visualized in the electron microscope. Although most or all of the 17,000 different macronuclear molecules share this common property of circle-forming ability, it was not known how many different inverted terminal repeats, or "neck" sequences, exist among the collection of macronuclear DNA molecules. Subsequent experiments now indicate that a single such sequence is shared by the various macronuclear DNA molecules.

Several approaches have been used to reveal the nature and location of the repeat sequences. Duplex neck DNA was isolated directly from single-stranded circles by digestion with the single strand specific nuclease S_1.† This S_1-resistant material (neck DNA) migrates as a single band in polyacrylamide gels with a size of 23 bases, and after denaturation it reassociates rapidly as a single component.† Thermal melting curves for native and renatured neck DNA are identical.† These results suggested that either a unique neck sequence exists or that there is a small family of neck sequences of 23 BP, all present in nearly the same amounts, that do not crossreact to form heterologous duplexes. I now report on direct sequence analysis of the terminal regions of macronuclear DNA strands and of isolated neck DNA.

MATERIALS AND METHODS

***Oxytricha* DNA.** *Oxytricha* culture, macronuclear isolation, and DNA extraction were previously described (3). Purified neck DNA was prepared by a modification of the procedure of Herrick and Wesley.† Macronuclear DNA (160 μg) in low salt was boiled for 20 min and then brought to 0.18 M NaCl in a total volume of 350 μl. After annealing for 5 min at 56°, the sample was cooled to room temperature and 350 μl of reaction mixture was added so that final concentrations were: 0.18 M NaCl, 30 mM sodium acetate (pH 4.6), 0.5 mM $ZnCl_2$, plus 250 units of nuclease S_1 (ref. 7; a gift of Larry Gold). A 1-hr incubation at 24° was followed by phenol and ether extractions. Neck DNA was purified by electrophoresis in 40 cm long, 20% polyacrylamide (Bio-Rad)/7 M urea (Schwartz-Mann) gels (8) and eluted as in ref. 9.

DNA Labeling. ^{32}P was transferred to the 5′ termini of na-

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U. S. C. §1734 solely to indicate this fact.

Abbreviation: BP, base pair.

* Current address: Division of Biology, California Institute of Technology, Pasadena, CA 91109.

† G. Herrick and R. D. Wesley, unpublished.

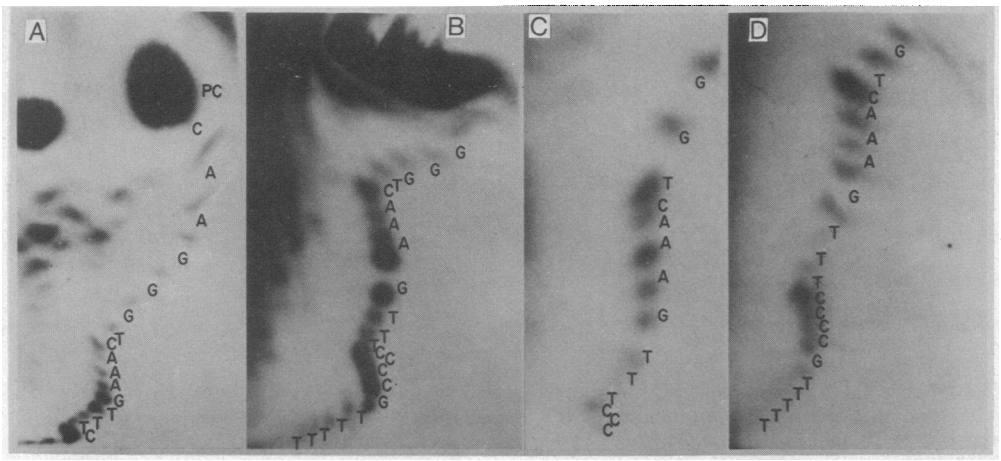


FIG. 1. Two-dimensional homochromatography of partial digests of 5′-^{32}P-labeled macronuclear DNA. First dimension is charge separation by electrophoresis at pH 3.5 on cellulose acetate. Size separation in the second dimension is by homochromatography on DEAE-cellulose using homomixes given below. (Homomixes are partially hydrolyzed solutions of yeast RNA.) (*A*) Homomix 2 (12). The 5′-terminal macronuclear sequence from residues 1–16 is shown. The initial nucleotide (pC) assignment comes from the complete digest experiment. (*B*) Homomix C (11). Crowding at the homomix front obscures residues 1–4, but a good trail of residues 5–26 is seen. (*C*) Homomix C. Although the early residues ran off the plate, this trail gives good resolution in the middle range, particularly the troublesome 5′-A-G-T-T-T series beginning at residue 12. That this is not alternatively A-G-G-G-G was confirmed by the failure of restriction enzyme *Hae* III to release a 16-base oligonucleotide from the end of macronuclear DNA. This enzyme recognizes the sequence G-G-C-C, which would be found in residues 15–18 in the alternate assignment scheme (14). (*D*) This homomix was prepared as in ref. 11, but was a 5% RNA solution with no alkaline hydrolysis. The DNA sample was alkaline hydrolyzed (*Materials and Methods*) to eliminate the possibility of any residual *Oxytricha* RNA. A defect in the plate or the homomix front obscured residues 1–6.

tive macronuclear DNA or neck DNA with T_4 polynucleotide kinase (New England Biolabs) and [γ-^{32}P]ATP (New England Nuclear) as described (8). Samples were freed of RNA by incubation with 100 μg of pancreatic RNase per ml (Worthington) in 0.15 M NaCl/15 mM sodium citrate at pH 7.0 for 1 hr at 37°. When noted, contaminating RNA was additionally hydrolyzed with 0.3 M NaOH for 1 hr at 37°, followed by neutralization and passage through a Sephadex G-50 column.

Sequence Analysis. 5′-End-labeled DNA was partially digested in 5- to 10-μl volumes in capillaries at 37°. Aliquots were withdrawn and frozen at 10, 20, 30, and 40 min. The reaction mixture contained about 1 μg of DNA resuspended in 50 mM Tris·HCl (pH 7.8)/5 mM $MgCl_2$/10 mM 2-mercaptoethanol/50 μg of bovine serum albumin per ml (Miles Laboratories) with 20 μg of yeast tRNA and 3 μg of sonicated calf thymus DNA as carriers, and 4 ng of DNase I (Worthington) and 1 μg of snake venom exonuclease (Worthington) as described (10). The aliquots were combined and analyzed by two-dimensional homochromatography on cellulose acetate strips (Schleicher and Schuell) and DEAE-cellulose plates (Polygram CEL 300 DEAE, Machery and Nagel and Co.) as described (10–12). Kodak NS-2T x-ray film was used for autoradiography. Complete digestion and 5′-terminal nucleotide analysis followed (13).

RESULTS

When the 5′ termini of total native *Oxytricha* macronuclear DNA were labeled with ^{32}P, 90% of the label was incorporated into 5′-dCMP (6). Two-dimensional homochromatography of 5′-end-labeled macronuclear DNA has now shown that these termini are identical beyond the first nucleotide. Sequencing trails (Fig. 1) produced with different homomixes indicate that the ends of most, if not all, of the thousands of different macronuclear DNA molecules contain the DNA sequence:

$^{5'}$pC-C-A-A-G-G-G-T-C-A-A-A-G-T-T-T-C-C-C-C-G-T-T-T-T-T
G-G-T-T-C-C-C-A-G-T-T-T-C-A-A-A-G-G-G-G-C-A-A-A-A-A

The common terminal sequence may extend for several nucleotides beyond the 26 BP shown, but further resolution has not been possible with these procedures. The left side of these chromatograms (Fig. 1) contains an amorphous background accounting for ≤50% of the label in various preparations. It is likely that nonspecific nuclease damage before labeling was the cause of much of this background. At the time of these experiments, no fresh macronuclear DNA could be prepared, and the sample that had once yielded 90% 5′-terminal dCMP had deteriorated to the point where only ~50% of the 5′ end label was incorporated into this nucleotide. However, in the light of previous results (6, †), I am confident that this homochromatography trail represents the major portion of the ends of the approximately 17,000 different macronuclear DNA molecules.

To test further the sequence homogeneity of the termini, I isolated duplex neck DNA from single-strand circles. Circles were formed by denaturing macronuclear DNA and allowing it to briefly reanneal. All but the duplex regions of this circular DNA was then digested with S_1 nuclease. The resulting neck DNA was purified by gel electrophoresis, 5′ ends were labeled, and the DNA was then subjected to two-dimensional homochromatography. The autoradiograph reveals a trail of spots representing a single nucleotide sequence (Fig. 2). The sequence determined from the neck trail is contained in that previously found by analysis of 5′-end-labeled native macronuclear DNA, as expected. The first two nucleotides (5′pC-C) of the native terminal sequence (Fig. 1) were not assignable in the chromatograph of neck DNA, but the remainder of the trail corresponds to that given above, up to the limit of resolution.

Note that the extent of the inverted terminal repeat (neck) sequence as determined by 5′ labeling and sequencing of intact macronuclear DNA (≥26 BP) is somewhat longer than that determined by Herrick and Wesley (23 BP)† from gel electrophoresis or than the sequence shown in Fig. 2. The true extent of the neck sequence must be the larger value. It is likely that "breathing" at the ends of DNA duplexes during S_1 digestion results in shortening of the necks. Indeed, small amounts of DNA pieces larger than 23 bases do occasionally occur in the

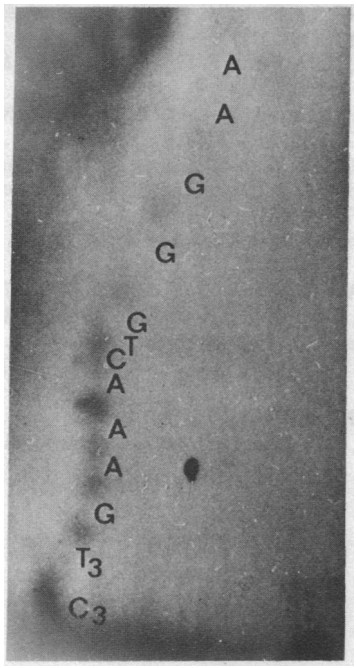


FIG. 2. Two-dimensional homochromatography of end-labeled neck DNA. Duplex neck DNA was prepared from circularized *Oxytricha* macronuclear DNA as described. The S_1-resistant material was 5′-^{32}P-labeled and purified from partial neck sequences (presumably due to "breathing" and digestion of the short duplex DNA) by electrophoresis in 20% polyacrylamide/7 m urea (8). The major band (about 23 bases) was eluted (9) and subjected to partial digestion and two-dimensional homochromatography as before. Nucleotide assignments are identical to those previously determined from 5′-terminal sequence analysis of whole macronuclear DNA, although assignment here begins at the third residue of the sequence presented in the *text*. The trail of the complementary strand was too obscured to be interpreted.

neck preparation gels of Herrick and Wesley (personal communication).

Finally, as additional evidence that the neck sequence is present at the ends of intact macronuclear molecules, the 5′ termini of macronuclear DNA were first labeled with ^{32}P, and a portion of this radioactivity recovered in neck DNA subsequently purified from this material. (This is a reversal of the order of the neck preparation and end-labeling procedure used in Fig. 2.)

In summary, several lines of evidence indicate that all of the thousands of different macronuclear DNA molecules share an identical inverted terminal, or neck-forming, sequence. All or nearly all of the DNA strands can form single-stranded circles held together by a short region at or near the ends. When these duplex regions are isolated as neck DNA, they are found to be homogeneous in size, and they reassociate extremely rapidly as a single component with no apparent mismatching. In addition, a major fraction of all the natural 5′ termini have the same sequence for at least 26 bases in from the end, and this same sequence is present in the necks.

DISCUSSION

Hypotrichous ciliates such as *Oxytricha* are of special interest because all the DNA in the macronucleus occurs naturally as "gene-sized" molecules. We have now demonstrated that after excision from precursor polytene chromosomes, these small DNAs are left with identical sequences of about 26 BP, arranged in reverse order, at each end of the molecules. In addition, comparison of the contour lengths of native and denatured

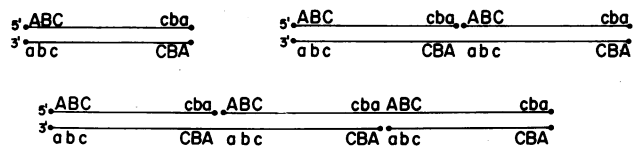


FIG. 3. A model of possible arrangements of neck sequence copies of native macronuclear DNA molecules. The neck ($\frac{ABC}{abc}$) sequences, corresponding to the 26-BP sequence determined, are shown at the ends of macronuclear DNA molecules and at internal locations at sites of single-strand interruptions. Every denatured single strand is capable of forming a circular structure closed by a duplex neck region, or a more complex reannealed structure. See ref. 6 for further discussion and G. Herrick and R. D. Wesley (unpublished). Note that if the internal double copy of the repeat sequence is necessary for DNA replication, as proposed by Heumann (18), molecules such as the first example shown would be prohibited.

macronuclear DNA, as well as quantitation of 5′-terminal nucleotide labeling, indicate that macronuclear DNA molecules of *Oxytricha* contain a number average of ~1.7 single-strand interruptions per molecule (5, 6, †). Since all experimental procedures involving the terminal or the neck sequences (e.g., circle formation, 5′-end labeling and sequencing) involved denatured DNA, the neck-forming sequences must also be present at the "internal" strand ends. We thus presume the existence of the types of macronuclear DNA molecules diagrammed in Fig. 3. The neck sequence exists both as an inverted terminal repeat and as a palindrome surrounding single-strand interruptions.

It is extraordinary that all, or nearly all, of the approximately 17,000 different macronuclear DNA molecules share this identical repeat sequence. At this point, one can only speculate upon its function. Inverted repeat sequences have been found in the chromosomes of many organisms, and several functions have been postulated for them (15–17). There are at least two times in the life cycle of *Oxytricha* when this sequence arrangement could prove useful. One is during the process of cleavage and sequence diminution by which the "gene-sized" molecules of macronuclear DNA arise from polytene chromosomes. The terminal copies of the neck sequence may serve as "punctuation" for this process. If this is the case, the cell must be able to distinguish between the terminal copies of this sequence and internal pairs surrounding single-strand interruptions (see Fig. 3). In addition, the repeat sequences might play a role in the replication of macronuclear DNA during vegetative cell growth. Heumann (18) has proposed a mechanism by which internal and terminal repeats of the neck sequence interact to allow replication of the ends of these small linear DNA molecules. This model requires a palindrome-specific nicking enzyme. Indeed, such an enzyme must be responsible for the single-strand interruptions that are introduced into *Oxytricha* macronuclear DNA during every cell cycle. Other functions for this sequence arrangement are of course possible, and multiple uses of the neck sequences should not be ruled out.

Further experiments are continuing to clarify the macronuclear gene structure. I have constructed recombinant DNA plasmids containing intact macronuclear DNA molecules to study features of these molecules on an individual basis (6). In addition to demonstrating other properties of *Oxytricha* DNA, the study of recombinant plasmids can show the relationship of the repeated sequences to the coding portion of the DNA in macronuclear molecules for which the appropriate message is available.

I thank David Prescott for his support of this research and assistance in preparing this manuscript, and Glenn Herrick and Ronald Wesley

for communication of unpublished results. I also thank Tom Maniatis for assistance in learning the DNA sequencing techniques. This work was supported by National Institutes of Health Grant GM19199 to D. Prescott.

1. Prescott, D. M., Murti, K. G. & Bostock, C. J. (1973) *Nature* **242**, 576–600.
2. Spear, B. B. & Lauth, M. R. (1976) *Chromosoma* **54**, 1–13.
3. Lauth, M. R., Spear, B. B., Heumann, J. & Prescott, D. M. (1976) *Cell* **7**, 67–74.
4. Prescott, D. M. & Murti, K. G. (1974) *Cold Spring Harbor Symp. Quant. Biol.* **38**, 609–618.
5. Wesley, R. D. (1975) *Proc. Natl. Acad. Sci. USA* **72**, 678–682.
6. Lawn, R. M., Heumann, J. M., Herrick, G. & Prescott, D. M. (1978) *Cold Spring Harbor Symp. Quant. Biol.* **42**, in press.
7. Shenk, T. E., Rhodes, C., Rigby, P. W. & Berg, P. (1975) *Proc. Natl. Acad. Sci. USA* **72**, 989–993.
8. Maxam, A. M. & Gilbert, W. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 560–564.
9. Efstratiadis, A., Kafatos, F. C., Maxam, A. M. & Maniatis, T. (1976) *Cell* **7**, 279–288.
10. Maniatis, T., Jeffrey, A. & Kleid, D. (1975) *Proc. Natl. Acad. Sci. USA* **72**, 1184–1188.
11. Barrell, B. G. (1971) in *Procedures in Nucleic Acids Research*, eds. Cantoni, C. L. & Davies, D. R. (Harper and Row, New York), Vol. 2, pp. 751–795.
12. Jay, E., Bambard, R., Padmanabagan, R. & Wu, R. (1974) *Nucleic Acids Res.* **1**, 331–353.
13. Bernardi, A. (1974) *Anal. Biochem.* **59**, 501–507.
14. Roberts, R. J. (1976) *CRC Crit. Rev. Biochem.* **4**, 123–164.
15. Cavalier-Smith, T. (1974) *Nature* **250**, 467–470.
16. Pellegrini, M., Manning, J. & Davidson, N. (1977) *Cell* **10**, 213–224.
17. Kleckner, N. (1977) *Cell* **11**, 11–23.
18. Heumann, J. M. (1976) *Nucleic Acids Res.* **3**, 3167–3171.